Prior listening in rooms improves speech intelligibility

Eugene Brandewie and Pavel Zahorik[a)]
Department of Psychological and Brain Sciences, Heuser Hearing Institute, University of Louisville, Louisville, Kentucky 40292

(Received 9 February 2009; revised 20 April 2010; accepted 4 May 2010)

Although results from previous studies have demonstrated that the acoustic effects of a single reflection are perceptually suppressed after repeated exposure to a particular configuration of source and reflection, the extent to which this dynamic echo suppression might generalize to speech understanding in room environments with multiple reflections and reverberation is largely unknown. Here speech intelligibility was measured using the coordinate response measure corpus both with and without prior listening exposure to a reverberant room environment, which was simulated using virtual auditory space techniques. Prior room listening exposure was manipulated by presenting either a two-sentence carrier phrase that preceded the target speech, or no carrier phrase within the room environment. Results from 14 listeners indicate that with prior room exposure, masked speech reception thresholds were on average 2.7 dB lower than thresholds without exposure, an improvement in intelligibility of over 18 percentage points on average. This effect, which is shown to be absent in anechoic space and greatly reduced under monaural listening conditions, demonstrates that prior binaural exposure to reverberant rooms can improve speech intelligibility, perhaps due to a process of perceptual adaptation to the acoustics of the listening room.

© *2010 Acoustical Society of America.* [DOI: 10.1121/1.3436565]

PACS number(s): 43.66.Lj, 43.71.Gv, 43.66.Pn [RLF] Pages: 291–299

I. INTRODUCTION

When sound propagates through a room, acoustic reflections from the room's surfaces interact with the source signal. When the source signal is speech, degradations in intelligibility can result. For example, it is well known that under listening conditions where the reverberation time is long (e.g., large spaces and/or hard reflecting surfaces), speech intelligibility is poor (Knudsen, 1929). Much of everyday speech communication, however, takes place in listening environments with shorter reverberation times. In these environments, speech intelligibility for normal-hearing listeners is often relatively unaffected, yet measurable physical distortions in the signal reaching the listener remain due to reflections from the room. Various perceptual mechanisms are likely responsible for this seeming insensitivity to acoustic reflections.

For sound localization, certain mechanisms that help to reduce the influence of early-arriving reflections have been extensively studied. The precedence effect (Wallach *et al.*, 1949) demonstrates the primacy of the first arriving waveform in specifying the apparent position of a sound source, and is often used to explain good localization capability in complex environments despite the presence of multiple acoustic reflections (see Litovsky *et al.*, 1999 for a comprehensive review of this literature). Neural correlates of the precedence effect have also been extensively studied. Evidence for the role of single neurons in the inferior colliculus in the spatial suppression of echoes has accumulated for a variety of species (Carney and Yin, 1989; Fitzpatrick *et al.*, 1995; Keller and Takahashi, 1996; Dent and Dooling, 2004). These results, in conjunction with most psychophysical evidence, suggest that the precedence effect is automatic, immediate, and likely subserved by relatively simple sub-cortical processes.

A similar type of echo suppression has been reported in the study of speech perception. Haas (1972) noted that a simulated acoustic reflection of a speech signal is not perceived as a separate sound source for delays ranging from approximately 10 to 30 ms. For this range of delays, the intensity of the reflection could also be increased by as much as 10 dB relative to the source before it was perceived as a separate sound source. Taken together, these results suggest a strong effect of echo suppression for intermediate delays. For longer delays, however, the reflection was perceived as a separate source and disturbed speech intelligibility results. It was therefore concluded that the length of delay between the lead and lag signals determines the salience of the reflection. This effect also appears to be relatively immediate and automatic.

More recent results challenge the view that echo-suppressive processes are automatic and immediate. Clifton has demonstrated that the strength of the echo suppression depends critically on recent stimulus history (Clifton, 1987; Clifton and Freyman, 1989). When listeners are presented with a repeating source signal (e.g., a train of clicks) along with a correspondingly repeating reflection (i.e., delayed copy of the source), the reflection becomes less audible over repetitions. This "buildup" of echo suppression can increase echo threshold (minimum delay resulting in a detectable echo) by roughly a factor of 2 when compared with standard precedence experiments using a single stimulus presentation (Freyman *et al.*, 1991) and can take tens of seconds to reach

[a)] Author to whom correspondence should be addressed. Electronic mail: pavel.zahorik@louisville.edu maximum suppression (Freyman *et al.*, 1991). Buildup has also been observed when two reflections are presented (Yost and Guzman, 1996), although this manipulation appears to affect the delays required to produce buildup effects when compared to single-reflection results. These results suggest that precedence buildup is therefore likely a form of unconscious perceptual adaptation based on repeated exposure to a given acoustic environment.

There is evidence to suggest that these adaptational aspects of the precedence effect may be mediated by more central brain processes than the known sub-cortical structures associated with standard, single presentation precedence effects. Left-right asymmetries in the buildup effect have been observed in which greater buildup in suppression was found when the reflection pulse-train was delivered on the listener's left side, than when on the right side, which is suggestive of cortical-level processing related to this effect (Grantham, 1996). Unilateral ablation of the auditory cortex in cats has also been shown to impair the echo suppression observed in the precedence effect for reflection locations ipsilateral to the lesion (Cranford *et al.*, 1971).

Precedence effect buildup can be dramatically destroyed when implausible or unnatural changes to the source and reflection relationship occur, such as an abrupt change in the spatial locations of the source and reflection (Clifton, 1987; Clifton and Freyman, 1989). This "breakdown" in precedence results in echo thresholds that are comparable to that observed for a single source+reflection stimulus presentation. It has been suggested that these dynamic aspects of the precedence effect may be indicative of processes that construct and neurally represent a model of the acoustic environment (Clifton *et al.*, 1994, 2002). Such a model would allow subsequent inputs to be evaluated in the context of the current acoustic environment, perhaps via a process of pattern classification (Blauert and Col, 1992), and could facilitate effective suppression of the potentially misleading spatial information resulting from reflections and reverberation. The buildup effect perhaps results from the experience-driven nature of the environmental models. Likewise, breakdown of this adaptation results when current sensory input becomes implausible in the context of the environmental model (Hartmann, 1997), and is manifest as a type of negative aftereffect. Such model-building processes do not appear to be mediated by cognition, however, since they have been shown to be resistant to practice and learning (Clifton and Freyman, 1997). Because dynamic echo suppression phenomena have most often been evaluated with only a very few simulated reflections (typically one), it is important to determine the extent to which these phenomena might generalize to more natural listening situations, such as reverberant rooms with multiple sound-reflecting surfaces.

Work by Djelani and Blauert (2001) has demonstrated that both buildup and breakdown phenomena do exist in situations with multiple reflections. In their experiment, a virtual triangular-shaped room environment was used to evaluate the effect of multiple reflections on echo thresholds using noise bursts. They used three preceding "conditioning" stimuli to build up echo suppression prior to the presentation of a test stimulus. The test stimulus was always in a particular triangular-shaped room. The preceding stimuli included an anechoic arrangement, a mirror-image of the triangle room used for testing, and the same room as the testing stimulus. Using an adaptive procedure, the authors scaled the size of the room to adjust the direct-to-reverberant energy ratio, and thus the intensity of the reflections relative to the source noise. The authors then measured the minimum room size required for the reflections to be barely audible. From this they found that trials using the same preceding room as the testing stimulus required a larger room size (greater direct-to-reverberant energy ratio) for detectable echoes, indicating that buildup to the room's reflections occurred and reduced the perceptual salience of the reflections. These results demonstrate that echo suppression can occur with multiple reflections, such as that found in everyday room environments.

In a separate experiment, Djelani and Blauert (2001) demonstrated that buildup and breakdown phenomena also exist when speech is used as a source signal. They allowed listeners to build up echo suppression to a continuous speech lead-lag stimulus (a repeated German sentence) in a loudspeaker arrangement similar to Freyman *et al.* (1991). The listeners would then press a button that altered the location of the lag stimulus: a manipulation known to cause a breakdown in echo suppression (Clifton, 1987). Following breakdown, listeners reported whether or not they perceived a "temporarily enhanced echo" (an increased salience of the lag stimulus). If there was no initial buildup to the continuous stimuli, the breakdown should have little effect on the audibility of the lag stimulus. Numerous "temporarily enhanced echoes" were reported, which indicated that a buildup of echo suppression had occurred using the continuous speech signal. The greatest enhancement occurred for delays between lead and lag pairs near echo threshold of the speech stimulus, which was determined in a separate experiment. Similar results were found with other types of source signals, thus demonstrating that dynamic echo suppression affects speech and other source signals in similar ways. Djelani and Blauert (2001) did not, however, examine a situation in which speech source signals were presented in a reverberant room environment with multiple sound reflecting surfaces. It is therefore not clear how buildup or breakdown might affect listening to speech in such complex environments.

Experiments by Watkins (2005a, 2005b) provide some clues to how speech perception may be affected in real room environments. Using a categorical perception task on a temporal-envelope continuum between "stir" and "sir" tokens, Watkins (2005a, 2005b) assessed the perceptual consequences of reverberant energy filling in the temporal gap following the stop consonant in "stir." A form of compensation for the effects of reverberant energy was identified when listening within a matching reverberant context provided by carrier phrases. For example, when the test word was presented in moderate reverberation without a matching reverberant context, listeners tended to more often report the target as "sir" than "stir." When a matching reverberant context was present, however, this shift was reversed when the level of reverberation in the carrier phrase context approached that of the target word. These experiments suggest that given the appropriate echoic context, a form of perceptual compensation, or echo suppression appears to lessen the deleterious effects of room reverberation on speech categorization. Presumably, this same effect would also result in improved speech intelligibility, although intelligibility was not explicitly measured in Watkins' (2005a, 2005b) studies.

Spatial separation of a speech source and a masking noise is also known to aid speech intelligibility. In anechoic space, the release from masking resulting from spatial separation can be as large as 9 dB (Hirsh, 1950), and is due largely to the acoustical shadow of the head causing an improved signal-to-noise ratio at one of the ears. Work by Plomp (1976) studied this spatial release from masking (SRM) effect in several reverberant spaces. In general, SRM was found to be inversely related to reverberation time. This effect may be explained by the fact that spatially diffuse reverberant energy causes the signal-to-noise ratios at the two ears to become more similar, and thus reduces the "better-ear advantage" that underlies SRM in anechoic space. Although not tested in Plomp's (1976) study, one might predict that prior exposure to the acoustics of the listening room environment might result in a perceptual suppression of the room's acoustical contributions, resulting in an improvement in speech intelligibility. This prediction is based both on the precedence effect buildup studies of Freyman *et al.* (1991) and Djelani and Blauert (2001), as well as the studies of Watkins (2005a, 2005b) that suggest a degree of suppression for the inherent acoustical contributions of the listening environment on the perception of speech sounds.

The current study focuses on this same prediction: That improvements to speech intelligibility within reverberant room environments will result from prior listening exposure in the same environment. The intention is to create an arrangement that allows for precedence effect buildup to potentially aid intelligibility by perceptually suppressing the acoustical contributions of the room. Where past studies have used trains of click stimuli to expose the listener to a particular spatial configuration of source and reflection (Freyman *et al.*, 1991; Clifton *et al.*, 1994), the current study uses preceding speech carrier phrases to provide listening exposure to a room environment, with multiple reflections and reverberation. Virtual auditory space techniques are used to simulate the acoustics of different rooms used in this study. These techniques allow for precise control of the acoustics of the listening configuration and the ability to easily switch between rooms from trial to trial in the experiments, a logistical impossibility with real room listening environments.

Three experiments were performed to study how exposure to a room environment might shape our perceptions of speech sounds. Experiment I measured closed-set speech intelligibility both with and without a preceding carrier phrase in a simulated reverberant environment. If the precedence buildup effect generalizes both to more natural room environments with multiple reflections and reverberation, and to speech intelligibility tasks, then the presence of preceding carrier phrases should result in improved intelligibility. Experiment II examined this same effect in anechoic space. If improvements in speech intelligibility with prior listening exposure are due only to suppression of the acoustical effects of the listening room, then little to no improvement should be observed when the room effects are absent (i.e., anechoic space). Experiment III tested the conditions of Exp. I monaurally. If sampling of the acoustical environment requires binaural input, prior exposure to the room acoustics will provide no benefit to speech intelligibility. These experiments are followed by a general discussion of the results.

II. EXPERIMENT I: PRIOR LISTENING EXPOSURE TO A REVERBERANT ROOM

A. Methods

1. Participants

Fourteen listeners (eight female) ages 17–24 years participated in the study. All had normal hearing as verified by audiometric screening at 20 dB HL at octave frequencies between 250 and 8000 Hz. Listeners were paid for their participation.

2. Stimuli

a. Room modeling Virtual acoustic techniques were used to simulate the room environments in this study. The techniques were identical to those described by Zahorik (2009), except that an equalization filter was applied to correct for the loudspeaker response used in the head-related transfer function (HRTF) measurement procedures. Briefly, this room simulation technique uses an image-model (Allen and Berkley, 1979) to precisely simulate early reflections and a statistical model to simulate late reverberant energy. The direct-path and early reflections are spatially rendered using HRTF measurements from a single participant (I.D. SXL). The result of the simulation is an estimated binaural room impulse response (BRIR) that describes the transformation of sound between the source and the listeners' ears in the simulated room. Overall, this simulation technique has been found to produce BRIRs that are reasonable physical and perceptual approximations to those measured in a real room (Zahorik, 2009).

Three rooms were simulated in this experiment (R1–R3), although psychophysical results were only analyzed from R2 (see Sec. II A 3 for design details). All had identical dimensions [length (x): 5.7 m; width (y): 4.3 m; height (z): 2.6 m], but varied in the absorptive properties of the reflecting surfaces. Within each simulated environment, the speech target was positioned 1.4 m directly in front of the listener. A spatially separated masker was presented on all trials and was positioned 1.4 m from the listener's position directly opposite the listener's right ear (90° azimuth angle). Figure 1 displays the dimensions of the simulated rooms, as well as target and masker positions.

Energy absorption coefficients (α) that control the absorptive properties of the reflecting surfaces in three simulated rooms are shown in Table I. Since the room simulation technique treats early reflections and late reverberation separately, there are separate sets of coefficients for each portion of the simulation, and only the late reverberation simulates any frequency-dependent absorption effects. The coefficients

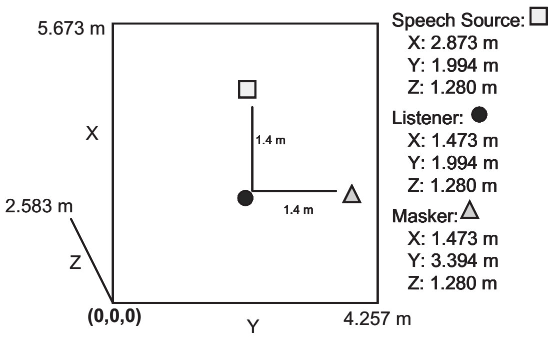


FIG. 1. Room dimensions and the position of the listener, the speech target, and the noise masker for all simulated rooms

for R2 were identical to those from Zahorik (2009), and were designed to approximate a real, moderately reverberant room (a large office room). R1 was less reverberant than R2, and R3 was more reverberant than R2. Reverberation properties for these rooms were controlled by adjusting the absorption coefficients by multiplicative factors of the coefficients from R2. Octave-band reverberation times, T_{60}, and clarity indices, C_{50}, are displayed for each simulated room in Table I. T_{60} is a measure of the time it takes sound to decay by 60 dB and C_{50} is a measure of the balance between early (0–50 ms) and late-arriving ($50-\infty$ ms) energy (ISO-3382, 1997). No attempt was made to equalize sound levels across the three rooms. Hence the more reverberant rooms produced greater at-the-ear sound levels than the less reverberant rooms.

b. Speech corpus and masker The speech stimuli used in this study were from the coordinate response measure (CRM) corpus (Bolia *et al.*, 2000). Each speech sentence in this corpus has the format "Ready ⟨Call Sign⟩ go to ⟨Color⟩ ⟨Number⟩ now." The corpus features eight talkers (four male and four female), eight call signs (Charlie, Ringo, Laker, Hopper, Arrow, Tiger, Eagle, Baron), four colors (Blue, Red, White, Green), and eight numbers (1–8). All combinations were used in this study.

Two conditions were created based on the length of the speech carrier phrase that preceded the target phrase, the idea being that longer carrier phrases would increase exposure to the acoustics of the room prior to the target. In the *sentence carrier* (SC) condition, two full-length CRM sentences were presented sequentially with approximately 2.5 s of silence between the sentences. The talker and call-sign for the first sentence was selected at random, but the second (target) sentence always had the same talker as the first sentence and always had the call-sign 'Baron'. In the *no carrier* (NC) condition, the target color and number were presented alone, without any carrier phrase (i.e., 'Green Three').

In both conditions, speech targets were presented in the presence of a spatially separated Gaussian noise masker (see Fig. 1). In the NC condition, the masker was 4 s in duration and preceded speech target onset by 1 s. In the SC condition, the masker was 10 s in duration and preceded the onset of the carrier phrase by 500 ms. Speech and masker signals were convolved with the BRIRs for their appropriate locations relative to the listener in either R1, R2, or R3 in order to simulate the spatial listening configuration shown in Fig. 1 for each room. Representative temporal waveforms for the speech target and the noise masker in R2 are displayed in Fig. 2 for both conditions in the experiment.

All stimuli were presented over equalized headphones (Beyerdynamic DT-990 Pro) at a moderate level (70 dB SPL peak at the entrance to the ipsilateral ear).

TABLE I. Energy absorption input parameters (early and late α) for the room simulation model and resulting room acoustic parameters (T_{60} and C_{50}) estimated from model BRIR outputs for each simulated room (R1–R3).

Room	Center frequency (Hz)	R1	R2	R3
Early alpha	Broadband	0.390	0.290	0.041
Late alpha	125	0.533	0.400	0.057
	250	0.400	0.300	0.043
	500	0.400	0.300	0.043
	1000	0.400	0.300	0.043
	2000	0.293	0.220	0.031
	4000	0.267	0.200	0.029
T_{60} (s)	Broadband	0.316	0.420	2.966
	125	0.364	0.447	2.655
	250	0.293	0.412	2.750
	500	0.328	0.429	2.748
	1000	0.368	0.487	3.015
	2000	0.317	0.444	3.430
	4000	0.096	0.135	1.581
C_{50} (dB)	Broadband	25.8	13.4	−6.6
	125	8.5	−6.0	−23.5
	250	11.1	0.5	−17.8
	500	11.1	−0.4	−19.1
	1000	8.9	−2.8	−21.5
	2000	27.5	14.5	−3.8
	4000	42.3	29.5	20.5

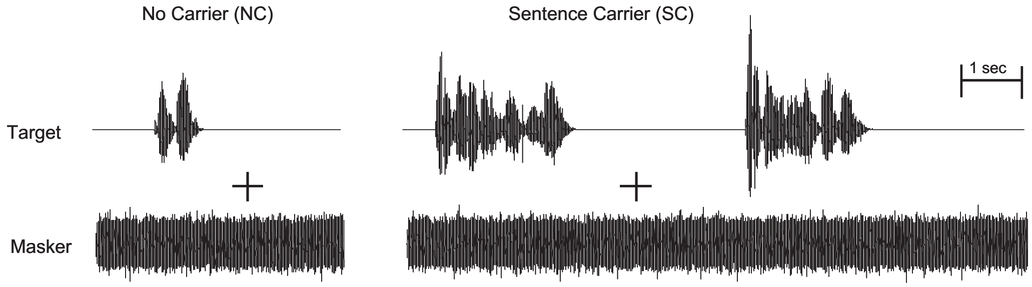


FIG. 2. Examples of the target and masker waveforms (left-ear only) for the NC and SC conditions in a moderately reverberant room (R2).

3. Design

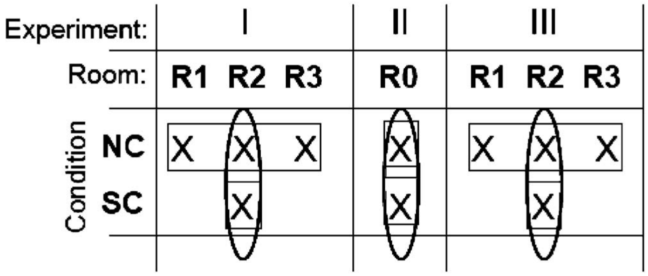
All listeners were tested in both the NC and SC conditions. In both conditions, stimuli were presented in blocks of 54 trials which contained nine signal-to-noise ratios (SNRs): −28 to +4 dB in 4 dB steps. SNR was manipulated by adjusting the gain of the speech target signal prior to convolution with the BRIRs. The masker level was fixed. Target color and number, and the SNR were selected at random for each trial.

In the NC condition blocks, the room environment was selected at random (equal probability) from trial to trial across the block from one of the three reverberant rooms (R1, R2, and R3). This manipulation was designed to minimize any carry-over effects from exposure to a particular room from trial to trial. Each SNR in each room was presented twice within a block of NC trials. In the SC condition, R2 was tested exclusively across the block in an attempt to maximize room exposure. Here each SNR was presented six times within a block of trials. To quantify any improvements in speech intelligibility, performance in the SC condition was compared to performance in R2 only of the NC condition. Results from R1 and R3 in the NC condition were not analyzed in this study. Table II further illustrates the block design in the experiment. Each listener completed 15 blocks of trials for the NC condition, and 5 blocks of trials for the SC condition. This yielded an equal number of responses (270 trials) in R2 for each condition.

4. Procedure

The listener was seated in a sound-attenuating chamber (Acoustic Systems, Austin, TX—custom double wall) and listened to the headphone-presented stimuli. The listener's task was to select the appropriate color and number combination using a computer mouse on a graphical interface. Feedback as to whether the response was correct was provided after every trial. All stimulus presentation and response collection was implemented using MATLAB software (Mathworks Inc., Natick, MA).

TABLE II. Diagram of trial blocking procedures for all conditions and experiments. An 'X' indicates a given combination of room and condition presented within a block of trials. The ellipses indicate conditions used in analyzing the room exposure effects of interest in this study (the difference of SC and NC).



	Experiment:	I			II	III		
	Room:	R1	R2	R3	R0	R1	R2	R3
Condition	NC	X	X	X	X	X	X	X
	SC		X		X		X	

5. Data analysis

For each listener, the proportion of correct color/number responses, P(C), was computed for R2 at all SNRs in both the NC and SC conditions. Similar P(C) computations were also made based on data pooled across all listeners. Logistic functions were then fit to the P(C) data using a maximum-likelihood algorithm (Wichmann and Hill, 2001b, 2001a) to approximate the psychometric function for each listener (or pooled data) in a given condition. These functions (Ψ) are defined by the following relationship:

$$\psi = (1-\delta) \times \frac{1}{1+\exp(-(x-a)/b)} + \delta. \qquad (1)$$

Delta (δ) is the lower asymptote of the function, which is set to the chance performance level of 1/32 (3.125%) in this task. a is the threshold parameter: the SNR corresponding to the midpoint of Ψ between perfect performance (100%) and chance performance (3.125%), which is 51.56%. b is a parameter which is inversely related to the slope of the function around its midpoint. 95% confidence intervals were estimated for fitted thresholds using a bootstrapping procedure (Wichmann and Hill, 2001b, 2001a). Goodness of fit was evaluated by noting the proportion of variance in the observed data accounted for by each function (R^2).

B. Results and discussion

In general, the proportion of variance accounted for by the fitted functions, R^2, was high, with a minimum of 0.94 and a median of 0.98. Visual inspection of the logistic fits revealed that the functions were quite homogeneous across listeners and conditions. Function slope parameters (b) were quite similar as well. The values of b in this analysis had a mean of 3.545 and a standard deviation of 0.142, which at the midpoints of the functions corresponded to a mean slope of approximately 0.0701 P(C)/dB, with a standard deviation of 0.0027 P(C)/dB. This high degree of similarity in function slopes suggests that these functions can be accurately described by their threshold (a) values alone. Figure 3 displays

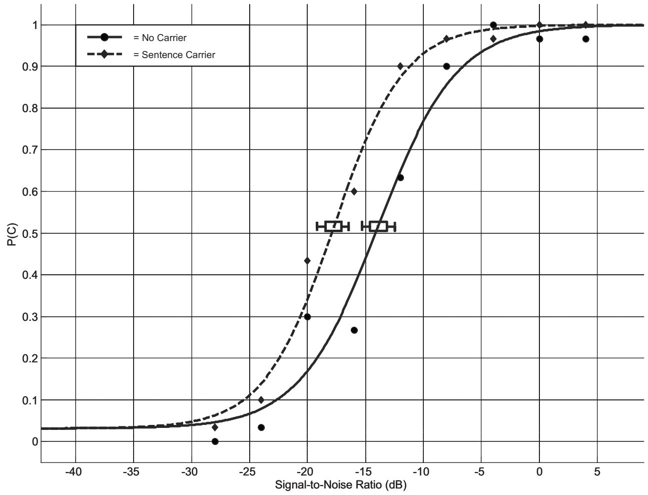


FIG. 3. Proportion of correct responses, P(C), as a function of signal-to-noise ratio for a single listener (LIS) in Exp. I. Data from the moderately reverberant room (R2) for both the NC and SC conditions are shown, along with logistic function fits for each condition (see text for details). Each data point is based on responses from 30 trials. Speech reception threshold estimates and their 95% confidence intervals are indicated for each curve at the midpoint between chance and perfect performance.

TABLE III. Room exposure effect size (NC threshold–SC threshold) in decibels, Δ dB, and the corresponding improvement in speech intelligibility (proportion correct) with room exposure measured at threshold SNR for the NC condition, Δ P(C), for all listeners in the R2 room environment.

	R2	
SUB	ΔdB	ΔP(C)
‘LJL’	1.47	7.9
‘LIF’	1.58	10.0
‘LJC’	1.65	11.0
‘LIB’	2.38	11.7
‘LJE’	1.96	15.0
‘LIX’	2.87	16.0
‘LJJ’	2.72	17.2
‘LJK’	1.98	20.1
‘LIK’	3.43	21.7
‘LJG’	2.69	21.7
‘LJI’	3.79	24.1
‘LJB’	3.44	24.4
‘LIT’	3.75	26.2
‘LIS’	3.82	27.1
Mean	2.68	18.14
Median	2.71	18.61

an example of the fitted psychometric functions for both conditions in R2 for a single listener. A decrease of approximately 3.8 dB in speech reception threshold for the SC condition relative to the NC condition is immediately apparent. Most listeners showed qualitatively similar results, but differed in the magnitude of the effect.

A summary of the threshold results from all listeners for R2 is shown in Fig. 4. The mean thresholds across listeners in R2 (n=14) was −13.50 dB (standard deviation σ=0.91) for the NC condition and −16.18 dB (σ=0.91) for the SC condition, yielding an average decrease in threshold of 2.68 dB. Similar results are observed in the thresholds based on data pooled across listeners (on the far right in Fig. 4). The effect was also consistent across listeners; all listeners demonstrated decreased threshold of at least 1.4 dB with prior exposure to R2. Results of a matched-sample t-test revealed a significant difference in thresholds between the two conditions in the R2 environment, $t(13)=11.618$, $p<0.0001$. These results clearly demonstrate that prior listening exposure to a reverberant room results in decreased speech reception thresholds.

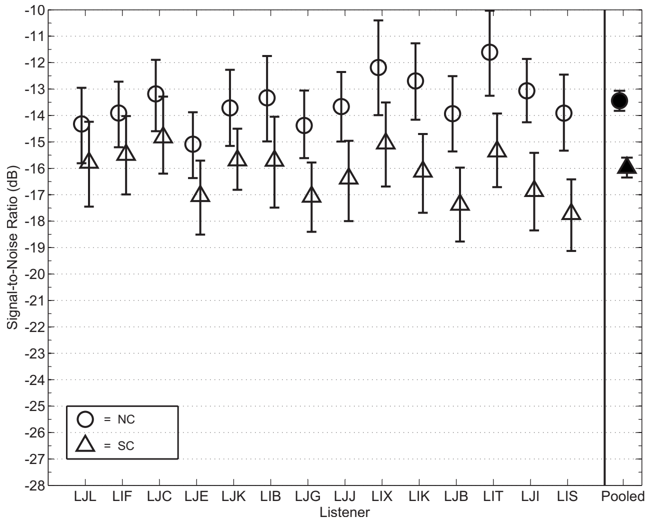


FIG. 4. Speech reception thresholds from Exp. I for the NC and SC conditions in room R2. 95% confidence intervals are displayed for each threshold estimate. Listeners are rank-ordered from left to right by effect size (NC threshold–SC threshold). Threshold values based on function fits to the data pooled across all listeners are shown on the right.

To further interpret the effect of room exposure on speech reception thresholds visible in Fig. 4, the effects were translated to equivalent improvements in the percentage of correctly identified speech targets. This translation was accomplished by evaluating the fitted psychometric function for the SC condition at the threshold SNR for the NC condition for each listener. The results of these translations are shown in Table III, which displays threshold difference (NC–SC) in the two conditions (dB) for each listener, paired with an equivalent change in speech intelligibility. From these results it is clear that even relatively small changes in threshold values between the two conditions can render substantial improvements in speech intelligibility (greater than 18 percentage points on average) in this listening situation.

In order to assess whether longer-term exposure to a room (over many trials within a block) could further improve speech intelligibility, an analysis of the data from SC condition (where the listening room was fixed within a block of trials) was performed. The analysis partitioned SC blocks into thirds, one with the first 18 trials in the block, one with the middle 18 trials, and one with the last 18 trials. Thresholds were then computed across all listeners for each third of the blocks. Results of a repeated-measures ANOVA revealed no significant change in thresholds, thus no improvement in speech intelligibility, across the partitions, $F(2,12)=0.742$, p=0.50. This suggests that after initial exposure and the resulting improvement in speech intelligibility that occurs

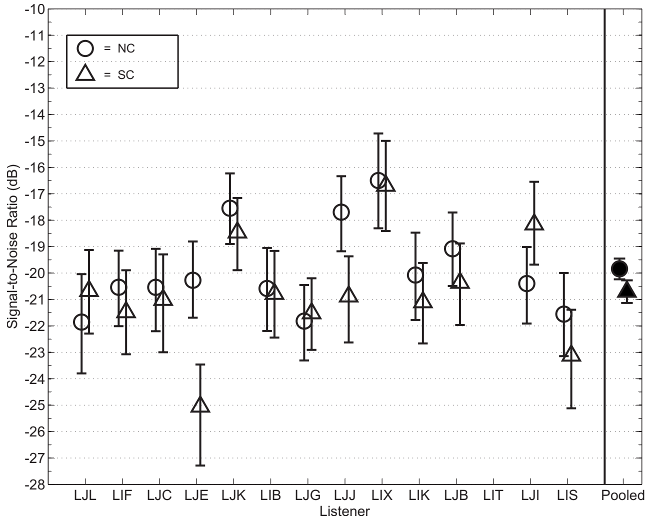


FIG. 5. Speech reception thresholds from Exp. II for the NC and SC conditions in anechoic space (R0). 95% confidence intervals are displayed for each threshold estimate. The left to right listener order is the same as shown in Fig. 4, except that listener LIT did not participate in Exp. II. Threshold values based on function fits to the data pooled across all listeners are shown on the right.

within a single trial, little to no additional improvement is observed across multiple trials. Hence, the processes underlying this room exposure effect appear to be quite fast-acting (on the order of seconds). Because this analysis was only able to examine room exposure times on the order of minutes, the extent to which much longer-term exposure effects also may affect speech intelligibility is not currently known. There is evidence to suggest that long-term exposure (on the order of many hours) to a reverberant room can affect sound localization performance (Shinn-Cunningham, 2000).

III. EXPERIMENT II: ANECHOIC SPACE

A. Methods

The stimulus generation methods used in Exp. II were identical to Exp. I, except that the simulated listening environment was an anechoic room (R0). The anechoic space had the same dimensions and source/listener configuration as R2 except the absorption coefficients were all set to unity (complete absorption). As a result, T_{60} values for this rooms were all very near zero, and C_{50} values were all greater than 100 dB.

The design of Exp. II was also identical to Exp. I, except that the NC condition no longer included trials from other simulated rooms. This was done because significant carry-over effects across trials (e.g., room exposure effects spanning multiple trials) were not observed in Exp. I. Hence, both NC and SC conditions in this experiment contained only R0. See Table II for an illustration of the block design in this experiment. Thirteen out of the fourteen listeners from Exp. I participated in Exp. II (all except LIT).

B. Results and discussion

Speech reception thresholds from the anechoic space tested in this experiment (R0) are displayed in Fig. 5, using analysis techniques identical to those implemented in Exp. I. Threshold estimates based on pooled data are shown on the far right of Fig. 5. Overall, thresholds from Exp. II were lower than those from Exp. I. This was due to the removal of the acoustic reflections and likely the resulting improvement in the effective SNR at the ear contralateral to the masker. Similar effects have been well-documented in previous literature (Plomp, 1976).

Much more relevant to the current study is the fact that in R0, threshold values were in most cases quite similar between the NC and SC conditions. The mean threshold ($n=13$) was -19.88 dB (standard deviation, $\sigma=1.70$) for the NC condition and -20.72 dB ($\sigma=2.12$) for the SC condition, which yielded an average effect size of only 0.84 dB. Although two listeners did have effect sizes considerably larger than the group average (LJE, 4.78 dB; LJJ, 3.19 dB), results of a matched-sample t-test confirmed that no statistically significant differences exist between the NC and SC condition in R0 for this group of listeners as a whole, $t(12)=1.701$, $p=0.11$. Overall, these results suggest that little to no improvement in speech intelligibility can be attributable to the presence of the carrier phrase itself for most listeners. Considering the results of both Exp. I and Exp. II, it may be concluded that the effect of prior room listening exposure on speech intelligibility appears to be specific to reverberant listening environments.

IV. EXPERIMENT III: MONAURAL PRESENTATION

A. Methods

The methods and participants in Exp. III were identical to Exp. I, except that all sound stimuli were presented monaurally. Monaural stimuli were generated by digitally removing the right-ear signals and retaining the left-ear signals (contralateral to the masker) from all stimuli used in Exp. I. Even though the results of Exp. I demonstrated no across-trial exposure effects, the NC condition in Exp. III was tested identically to Exp. I, with R1, R2, and R3 all presented within a block of trials in randomized order. As in Exp. I, the SC condition contained stimuli only from R2, and all analyses concerned only results from R2. The block design for Exp. III is shown Table II. All 14 listeners from Exp. I participated in Exp. III.

B. Results and discussion

A summary of the speech reception thresholds from the monaural presentation of R2 is displayed in Fig. 6, using analysis techniques identical to Exp. I. Overall, monaural thresholds in both NC and SC conditions were elevated relative to the comparable binaural thresholds reported in Exp. I. This effect is almost surely related to the well-known binaural masking release, and effects of similar magnitude have been reported in previous studies conducted under comparable conditions in reverberant listening environments (Plomp, 1976).

More import in the context of this study is that the effect of prior room listening exposure demonstrated in Exp. I appears to be greatly reduced. As is apparent from the results displayed in Fig. 6, most listeners show little difference in threshold estimates between the NC and SC conditions. The

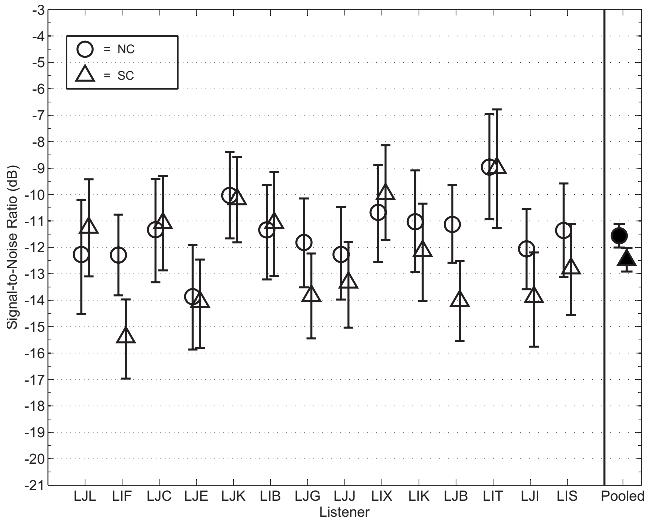


FIG. 6. Speech reception thresholds from Exp. III for the NC and SC conditions in R2 under monaural listening conditions (left-ear only). 95% confidence intervals are displayed for each threshold estimate. The left to right listener order is the same as shown in Fig. 4. Threshold values based on function fits to the data pooled across all listeners are shown on the right.

mean threshold ($n = 14$) from monaural presentation was -11.46 dB (standard deviation, $\sigma = 1.16$) for NC and -12.29 dB ($\sigma = 1.90$) for SC, yielding an average effect size of just 0.83 dB. Two listeners, LJB and LIF, exhibited relatively large effect sizes, however. Including these two listeners in a matched-pair *t*-test on the differences in thresholds between NC and SC conditions did yield a statistically significant decrease in thresholds, $t(13) = 2.38$, $p = 0.03$, although the magnitude of the effect is still much smaller than that observed under binaural listening conditions (Exp. I). If these two listeners are excluded from the analysis, then no significant difference between the NC and SC conditions is observed for monaural listening, $t(11) = 1.64$, $p = 0.13$. Although the source of this individual variability is not known, one possibility is that the room adaptation effect observed in this study may depend on a combination of both monaural and binaural information and that listeners may differ in their relative weighting of these two sources of information. Such a difference in weighting strategies would not be unlike that observed in other complex perceptual tasks in audition such as spectral profile analysis (Berg and Green, 1990) and sound localization (Macpherson and Middlebrooks, 2002; Zahorik, 2002). For most listeners, however, monaural input is not sufficient to produce significant improvements in speech intelligibility with prior listening exposure. It may therefore be concluded the room exposure effects demonstrated in Exp. I are primarily binaural phenomena.

V. GENERAL DISCUSSION

Results from this study demonstrate a significant increase in speech intelligibility (decreased speech reception threshold) in a condition that provided prior listening exposure to a reverberant room relative to one that did not (Exp. I). This effect is relatively absent in the anechoic space (Exp. II) and seems to rely on binaural input for most listeners (Exp. III). This suggests that increased binaural exposure to the reverberant room environment (and not simply the anechoic speech signal) may allow for a mechanism of perceptual adaptation to adjust to the unique acoustical characteristics of the room environment in an attempt to maximize speech intelligibility.

Although these results are not the first to demonstrate a precedence-like buildup effect with continuous speech (see Djelani and Blauert, 2001), to our knowledge, they are the first to demonstrate an improvement in speech intelligibility in a realistic room environment with multiple reflecting surfaces. Most studies in this area have examined only the effect of a single reflection on a target source. These single-reflection paradigms have provided insights into the dynamics of the precedence effect, but until now, it has been uncertain whether such effects generalize to more complex, real room environments.

One reason to suspect that the results reported in this study are related to precedence effect buildup is that both appear to involve echo suppression processes that do not operate immediately, but instead take some degree of time to accumulate and adapt (hence the "buildup" moniker). Although results from the current study suggest that significant buildup to speech in reverberant rooms exists after just a few seconds (i.e., a two-sentence carrier phrase) compared to situations with no prior listening exposure, the precise time-course of this effect is clearly an area in need of further study.

It is also interesting to speculate about the potential for additional uncontrolled buildup in the NC condition in the current study, given that the onset of the noise masker preceded the speech target by 1 s. If true, the effect of prior listening exposure to the room may have been even greater had less masker/target onset asynchrony been used. During informal pilot testing for this study, simultaneous onsets for the NC condition were initially tried, but later discarded because listeners reported having very poor target intelligibility in this case. Perhaps this poor intelligibility was due to a distinct lack of buildup. Formal testing will be needed to explore this possibility and to determine more generally the relative contributions of the speech carrier phrases and noise maskers in facilitating adaptation to room acoustics.

A second reason to suppose that the reported room exposure effects are similar to precedence effect buildup is that both appear to depend at least primarily on binaural input. Although many of the tasks used in the study of the precedence effect require spatial hearing proclivities that only the binaural system can provide, speech perception is perhaps primarily subserved by the monaural auditory system. It is clear, however, that many speech perception applications benefit from binaural input, such when speech targets are embedded in backgrounds of one or more competing, but spatially separated sources. Given the spatial configuration of target and masker in the current study, it is perhaps not so surprising that binaural system appears to play a key role in de-reverberating the room in this situation. Such a result does not imply that other monaural aspects of speech de-reverberation do not also exist, however. The perceptual compensation for room reverberation described by Watkins (2005a, 2005b) appears to operate with nearly equal strength in both monaural and binaural conditions, and perhaps functionally removes spectral colorations caused by room acoustics. This result appears fundamentally similar to observations described by Toole (2006) regarding loudspeaker reproduction in rooms, where listeners were found to be insensitive to the spectral characteristics of the room following room listening exposure. It therefore seems plausible to suppose that there may be separate and perhaps complementary aspects of room de-reverberation: one that relates to spatial configurations within the room and therefore is facilitated by binaural input, and one that is concerned primarily with removing monaural coloration caused by room acoustics. Such a two system hypothesis might also explain some of the sources of individual variability observed in the current study, particularly under monaural listening conditions. Future study will be needed to more fully test this hypothesis, and begin to address other important questions related to this effect such as the dependence of the effect on particular acoustical attributes of the room and the signals reaching the two ears.

VI. CONCLUSIONS

Speech intelligibility in a reverberant room improves with prior exposure to the acoustics of the room (18 percentage-point improvement on average compared to no exposure). This effect was absent in anechoic space and under monaural listening conditions for most listeners, and may result from a type of perceptual de-reverberation of the room environment. These results are consistent with the view that the physical effects of acoustic reflections may be suppressed via high-level perceptual processes that require adaptation time to the particular reflective listening environment in order to be effective.

ACKNOWLEDGMENTS

The authors wish to thank Noah Jacobs and Jeremy Schepers for their assistance in data collection. The authors also gratefully acknowledge the helpful comments of Dr. Rachel Keen (formerly Rachel K. Clifton), Dr. D. Wesley Grantham, an anonymous reviewer, and Dr. Richard L. Freyman (Associate Editor). Work supported by NIH-NIDCD (Grant No. R01DC008168).

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950.

Berg, B. G., and Green, D. M. (**1990**). "Spectral weights in profile listening," J. Acoust. Soc. Am. **88**, 758–766.

Blauert, J., and Col, J. P. (**1992**). "Irregularities in the precedence effect," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 531–538.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Carney, L. H., and Yin, T. C. T. (**1989**). "Responses of low-frequency cells in the inferior colliculus to interaural time differences of clicks: Excitatory and inhibitory components," J. Neurophysiol. **62**, 144–161.

Clifton, R. K. (**1987**). "Breakdown of echo suppression in the precedence effect," J. Acoust. Soc. Am. **82**, 1834–1835.

Clifton, R. K., and Freyman, F. L. (**1989**). "Effect of click rate and delay on breakdown of the precedence effect," Percept. Psychophys. **46**, 139–145.

Clifton, R. K., and Freyman, R. L. (**1997**). "The precedence effect: Beyond echo suppression," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), pp. 233–255.

Clifton, R. K., Freyman, R. L., Litovsky, R. Y., and McCall, D. (**1994**). "Listeners' expectations about echoes can raise or lower echo threshold," J. Acoust. Soc. Am. **95**, 1525–1533.

Clifton, R. K., Freyman, R. L., and Meo, J. (**2002**). "What the precedence effect tells us about room acoustics," Percept. Psychophys. **64**, 180–188.

Cranford, J. L., Ravizza, R., Diamond, I. T., and Whitfield, I. C. (**1971**). "Unilateral ablation of the auditory cortex in the cat impairs complex sound localization," Science **172**, 286–288.

Dent, M. L., and Dooling, R. J. (**2004**). "The precedence effect in three species of birds (Melopsittacus undulatus, Serinus canaria, and Taeniopygia guttata)," J. Comp. Psychol. **118**, 325–331.

Djelani, T., and Blauert, J. (**2001**). "Investigations into the build-up and breakdown of the precedence effect," Acta. Acust. Acust. **87**, 253–261.

Fitzpatrick, D. C., Kuwada, S., Batra, R., and Trahiotis, C. (**1995**). "Neural responses to simple, simulated echoes in the auditory brainstem of the unanesthetized rabbit," J. Neurophysiol. **74**, 2469–2486.

Freyman, R. L., Clifton, R. K., and Litovsky, R. Y. (**1991**). "Dynamic processes in the precedence effect," J. Acoust. Soc. Am. **90**, 874–884.

Grantham, D. W. (**1996**). "Left-right asymmetry in the buildup of echo suppression in normal-hearing adults," J. Acoust. Soc. Am. **99**, 1118–1123.

Haas, H. (**1972**). "The influence of a single echo on the audibility of speech," J. Audio Eng. Soc. **20**, 146–159.

Hartmann, W. M. (**1997**). "Listening in a room and the precedence effect," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and T. R. Anderson (Erlbaum, Mahwah, NJ), pp. 191–210.

Hirsh, I. J. (**1950**). "The relation between localization and intelligibility," J. Acoust. Soc. Am. **22**, 196–200.

ISO-3382 (**1997**)."Acoustics—Measurement of the reverberation time of rooms with reference to other acoustical parameters," International Organization for Standardization, Geneva.

Keller, C. H., and Takahashi, T. T. (**1996**). "Responses to simulated echoes by neurons in the barn owl's auditory space map," J. Comp. Physiol. **178**, 499–512.

Knudsen, V. O. (**1929**). "The hearing of speech in auditoriums," J. Acoust. Soc. Am. **1**, 56–82.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Macpherson, E. A., and Middlebrooks, J. C. (**2002**). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," J. Acoust. Soc. Am. **111**, 2219–2236.

Plomp, R. (**1976**). "Binaural and monaural speech-intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise)," Acustica **34**, 201–211.

Shinn-Cunningham, B. G. (**2000**). "Learning reverberation: Considerations for spatial audio displays," in 2000 International Conference on Auditory Display, Atlanta, GA.

Toole, F. E. (**2006**). "Loudspeakers and rooms for sound reproduction—A scientific review," J. Audio Eng. Soc. **54**, 451–476.

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (**1949**). "The precedence effect in sound localization," Am. J. Psychol. **62**, 315–336.

Watkins, A. J. (**2005a**). "Perceptual compensation for effects of echo and of reverberation on speech identification," Acta. Acust. Acust. **91**, 892–901.

Watkins, A. J. (**2005b**). "Perceptual compensation for effects of reverberation in speech identification," J. Acoust. Soc. Am. **118**, 249–262.

Wichmann, F. A., and Hill, N. J. (**2001a**). "The psychometric function: I. Fitting, sampling, and goodness of fit," Percept. Psychophys. **63**, 1293–1313.

Wichmann, F. A., and Hill, N. J. (**2001b**). "The psychometric function: II. Bootstrap-based confidence intervals and sampling," Percept. Psychophys. **63**, 1314–1329.

Yost, W. A., and Guzman, S. J. (**1996**). "Auditory processing of sound sources: Is there an echo in here?," Curr. Dir. Psychol. Sci. **5**, 125–131.

Zahorik, P. (**2002**). "Assessing auditory distance perception using virtual acoustics," J. Acoust. Soc. Am. **111**, 1832–1846.

Zahorik, P. (**2009**). "Perceptually relevant parameters for virtual listening simulation of small room acoustics," J. Acoust. Soc. Am. **126**, 776–791.